AD-A092 534 NORTH CAROLINA UNIV AT CHAPEL HILL DEPT OF STATISTICS F/G 12/1
POWER TRANSFORMATIONS WHEN THE CHOICE OF POWER IS RESTRICTED TO--ETC(U)
1980 R J CARROLL AFOSR-80-0080

UNCLASSIFIED AFOSR-TR-80-1052 NL

1 OF 1
AD
A092534

END
DATE
FILMED
1 81
DTIC

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered) UNCLASSIFIED

LEVEL II

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: AFOSR-TR-80-1052	2. GOVT ACCESSION NO. AD-A092534
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle): POWER TRANSFORMATIONS WHEN THE CHOICE OF POWER IS RESTRICTED TO A FINITE SET	5. TYPE OF REPORT & PERIOD COVERED: Interim rept.
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Raymond J. Carroll	8. CONTRACT OR GRANT NUMBER(s): AFOSR-80-0080
9. PERFORMING ORGANIZATION NAME AND ADDRESS: University of North Carolina, Department of Statistics, Chapel Hill, North Carolina 27514	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 61102F 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS: Air Force Office of Scientific Research/NM, Bolling AFB, Washington, DC 20332	12. REPORT DATE: 1980
	13. NUMBER OF PAGES: 15
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): UNCLASSIFIED
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Power Transformations, Prediction, Box-Cox Family, Asymptotic Theory, Contiguity

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

We study the family of power transformatios proposed by Box and Cox (1964) whent he choice of the power parameter is restricted to a finite set Ω_R. The two cases in which obvious answers obtain are when the true parameter is an element of Ω_R and when λ is "far" from Ω_R. We study the case in which λ is "close" to Ω_R, finding that the resulting methods can be very different from unrestricted maximum likelihood and that inference depends on the design, the values of the regression parameters, and the distance of λ to Ω_R.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

182850

AD A092534

DDC FILE COPY

AFOSR-TR- 80 - 1052

Power Transformations when the Choice of Power is Restricted to a Finite Set

Raymond J. Carroll

University of North Carolina at Chapel Hill

Key Words and Phrases: Power transformations, prediction, Box-Cox family, asymptotic theory, contiguity.

AMS 1970 Subject Classifications: Primary 62 E 20; Secondary 62 G 35.

Research supported by the US Air Force Office of Scientific Research under contract AFOSR-80-0080.

Abstract

We study the family of power transformations proposed by Box and Cox (1964) when the choice of the power parameter λ is restricted to a finite set Ω_R. The two cases in which obvious answers obtain are when the true parameter λ is an element of Ω_R and when λ is "far" from Ω_R. We study the case in which λ_o is "close" to Ω_R, finding that the resulting methods can be very different from unrestricted maximum likelihood and that inference depends on the design, the values of the regression parameters, and the distance of λ to Ω_R.

1. Introduction

Box and Cox (1964) suggested the power family of transformations, wherein for some unknown λ ,

$$(1.1) \qquad y_i^{(\lambda)} = x_i\beta + \sigma\epsilon_i = \tau_i + \sigma\epsilon_i \qquad , \quad i=1,\dots,N .$$

Here the design vectors $x_i = (1\ c_{i2}\ \dots\ c_{ip})$, $\beta = (\beta_o\ \dots\ \beta_{p-1})'$, the ϵ_i are independent and identically distributed with mean zero, variance one and distribution F , and

$$y^{(\lambda)} = (y^{(\lambda)} - 1)/\lambda \qquad \lambda \neq 0$$
$$= \log y \qquad \lambda = 0 .$$

They studied both maximum likelihood and Bayes inference when F is the normal distribution. There is now a substantial literature on the problem, an incomplete list of which includes Andrews (1971), Atkinson (1973), Hinkley (1975), Bickel and Doksum (1980, denoted B-D), Carroll (1980), and Carroll and Ruppert (1980, denoted C-R).

B-D developed an asymptotic theory for estimation. If the normal theory MLE is $\hat{\beta}$ when λ is known and $\beta^* = \hat{\beta}(\hat{\lambda})$ when λ is unknown and estimated by $\hat{\lambda}$, they compute the asymptotic distributions of $N^{1/2}(\hat{\beta} - \beta)/\sigma$ and $N^{1/2}(\beta^* - \beta)/\sigma$ as $N \to \infty$, $\sigma \to 0$. These distributions are different, with the latter having a covariance matrix at least as large and often very much larger than that of the former; the estimates $\hat{\lambda}$ and β^* are highly variable and highly correlated in general. This suggests that there is a large "cost" due to estimating the power parameter λ . Unfortunately, these results (and independent Monte-Carlo work by Carroll (1980)) suggest that unconditional inference concerning β can be very difficult for, except in certain balanced designs, inference without taking into account the variability of $\hat{\lambda}$ will be incorrect while β^* is itself too variable to be much help. A theory for conditional inference might prove useful.

It is relevant to note that when $\beta=0$ and $\sigma=1$ are known, the curvature (Efron (1975)) for λ at $\lambda=0$ is $\gamma_o^2 = 10.67$; Efron suggests that a value $\gamma_o^2 \geq 1/8$ is "large"!

C-R study the prediction problem in the sense of estimating the conditional median of Y given a design point x_o ; this is inference in the original scale of the data. Their results are much more encouraging; while there is a cost due to estimating λ , it is generally not severe. For example, if $\beta = (\beta_o\ \beta_1\ \ldots\ \beta_{p-1})'$, the cost averaged over the distribution of the design is $(1+p)^{-1}$ (asymptotically as $N \to \infty$ and $\sigma \to 0$).

We are concerned with the following point which has been raised concerning the applicability of the B-D and C-R theories. In practice, one may be uncomfortable using an estimate such as $\hat{\lambda} = .037$, then the much more common log scale $(\hat{\lambda}=0)$ is "just as good". Thus it is reasonable to restrict the estimate of λ to a finite set Ω_R and to study the consequences of such a decision. Asymptotically, as $N \to \infty$ but λ and Ω_R stay fixed, one has the trivial results that if $\lambda \in \Omega_R$ one is almost always in the right scale so there are no difficulties, while if $\lambda \notin \Omega_R$ bias dominates and no useful results are obtainable.

In Table 1 we present the results of a Monte-Carlo study for estimating the conditional median of Y given x_o . The model is simple linear regression based on a uniform design with $\beta_o=5$, $\beta_1=2$ and

$$(1.2) \qquad N^{-1} \sum_{i=1}^{N} x_i' x_i = I_2 .$$

The errors were normally distributed with mean 0 and variance σ^2, and there were 500 replications of the experiment. The restricted power set was $\Omega_R = \{0, \pm 1/2, \pm 1\}$, and we made decisions in this set on the basis of the likelihood. For a given $\hat{\lambda}$, our estimator is

$$(1.3) \qquad \begin{array}{ll} (1 + \hat{\lambda}\, x_o\, \beta^*)^{1/\hat{\lambda}} & (\hat{\lambda} \neq 0) \\ \exp(\, x_o\, \beta^*) & (\hat{\lambda} = 0) \end{array}$$

The numbers listed in the Table 1 are the "relative mean square errors (MSE)", i.e., the mean square error of (1.3) divided by the MSE when λ is known. We list results for the origen $x_o = (1\ o)$ and when x_o is a randomly chosen number of the design; the latter is in effect an average relative MSE over the distribution of the design.

In Table 1 we see that the restricted estimator (RE) dominates the MLE when $\lambda=0$ (hence $\lambda \in \Omega_R$), while the MLE dominates when $\lambda \notin \Omega_R$. In this latter case note that increasing N or decreasing σ results in improved performance of the MLE relative to the RE.

In Table 2 we repeat the above experiment with the changes $\beta_o=7$, $\beta_1=4$. The slightly worse behavior of the MLE relative to the λ-known case is expected from the C-R theory. Note here that the change in parameter values causes the RE to be much worse than the MLE if $\lambda \notin \Omega_R$. Also, the effect of changing N or σ is highlighted.

From the Monte-Carlo, we see that the performance of the RE relative to the MLE depends on λ, N, σ and β. One purpose of the rest of this paper is to propose and investigate a simple theory which gives a somewhat more systematic understanding of this performance. More generally, we also investigate the question of the feasibility of constructing procedures for which the choice of λ is restricted but which also give performance comparable to the MLE.

Table 1

The MSE behavior of the MLE and RE relative to the λ-known estimate of the conditional median of Y given x_o. Here $\Omega_R = \{0, \pm 1/2, \pm 1\}$, $\beta_o = 5$ and $\beta_1 = 2$.

			ORIGIN (x_o = (0 1))			AVERAGE (x_o random member of design)		
N	σ	λ	MLE	RE	Ratio (High/Low)	MLE	RE	RATIO
20	1	0	1.17	1.05	1.11	1.07	1.01	1.06
20	1/2		1.28	1.18	1.08	1.17	1.01	1.16
40	1		1.13	1.00	1.13	1.08	1.00	1.08
40	1/2	0	1.22	1.00	1.22	1.16	1.00	1.16
20	1	1/8	1.13	1.26	1.12	1.07	1.15	1.07
20	1/2		1.25	1.43	1.14	1.13	1.24	1.10
40	1		1.12	1.29	1.15	1.03	1.10	1.07
40	1/2	1/8	1.25	1.53	1.22	1.05	1.24	1.18
20	1	1/4	1.12	1.17	1.04	1.06	1.10	1.04
20	1/2		1.24	1.34	1.08	1.12	1.24	1.11
40	1		1.10	1.29	1.17	1.04	1.10	1.06
40	1/2	1/4	1.23	1.52	1.24	1.06	1.14	1.08

Table 2

The MSE behavior of the MLE and RE relative to the λ-known estimate of the conditional median of Y given x_o. Here $\Omega_R = \{0, \pm 1/2, \pm 1\}$, $\beta_o = 7$ and $\beta_1 = 4$.

			ORIGEN (x_o = (0 1))			AVERAGE (x_o random member of design)		
N	σ	λ	MLE	RE	Ratio (High/Low)	MLE	RE	RATIO
20	1	0	1.34	1.00	1.34	1.25	1.00	1.25
20	1/2		1.43	1.00	1.43	1.27	1.00	1.27
40	1		1.23	1.00	1.23	1.32	1.00	1.32
40	1/2	0	1.37	1.00	1.37	1.27	1.00	1.27
20	1	1/8	1.25	1.72	1.38	1.15	1.61	1.40
20	1/2		1.37	2.24	1.64	1.18	2.30	1.95
40	1		1.24	1.80	1.45	1.04	1.59	1.53
40	1/2	1/8	1.38	2.47	1.79	1.07	2.33	2.18
20	1	1/4	1.24	1.65	1.33	1.13	1.49	1.32
20	1/2		1.36	2.52	1.85	1.16	2.19	1.89
40	1		1.22	2.11	1.73	1.05	1.42	1.35
40	1/2	1/4	1.36	3.41	2.51	1.09	2.20	2.02

2. A large sample theory

Any reasonable theory must have λ "close" to Ω_R for large sample sizes. We choose to do this by letting the cardinality of Ω_R increase with increasing sample size N and by letting $\lambda = \lambda_N$ converge to a fixed element of Ω_R. For ease of calculation we focus on the important special case that the log scale is "almost" correct, i.e., Ω_R always contains zero and

$$\lambda = b\sigma/N^{1/2} \quad .$$

Of course, when $b = 0$ the data truely have a log-normal distribution.

Let $\hat{\lambda}_R$ and $\hat{\lambda}_M$ denote the restricted and ML estimates of λ, let $\hat{\beta}_R$ or β^* be the estimate of β having chosen the power $\hat{\lambda}_R$ or $\hat{\lambda}_M$, and let

$$\begin{aligned} f(\lambda, x_o\beta) &= (1 + \lambda\, x_o\, \beta)^{1/\lambda} & (\lambda \neq 0) \\ &= \exp\,(x_o\, \beta) & (\lambda = 0)\ , \end{aligned}$$

which is the conditionional median of Y given x_o, with estimate (1.2). We assume the errors are normally distributed. Letting $e = (1\ o\ \ldots\ o)$, we assume

$$\underline{x}_i\, e' = 1 \qquad \text{(there is an intercept)}$$

$$N^{-1} \sum x_i' x_i = I \ .$$

Then, for any value of b, when λ is known the limit MSE is

$$\text{(2.2)} \qquad \mathrm{MSE}(\lambda \text{ known}) = \|x_o\|^2 \exp(2\, x_o\, \beta) \ .$$

For fixed σ the computations are very difficult, so we will follow the lead of Bickel and Doksum and consider only the case that $\sigma = \Gamma\,\eta$, where $\Gamma = \Gamma(N) \to 0$ is a known sequence; it simplifies notation to make the convention $\eta = 1$.

We are now in a position to define the restricted estimate $\hat{\lambda}_R$ of λ, which we take by convention to satisfy $|\hat{\lambda}_R| \leq 1$. Let $\mathcal{D} = \{d_k\}$ be a finite or countably infinite subset of the extended real line with $d_o = 0$, $d_{-k} = -d_k$, and $\sup\{d_k\} = \infty$. Define intervals midway between these points:

$$B_k = \left[(d_{k-1} + d_k)/2\ ,\ (d_k + d_{k+1})/2\right) .$$

Our restricted estimate $\hat{\lambda}_R$ satisfies $|\hat{\lambda}_R| \leq 1$ and maximizes the likelihood over the admissible set with $N^{1/2}\,\hat{\lambda}_R/\Gamma \in \mathcal{D}$. Asymptotically, the procedure becomes

> Choose $N^{1/2}\hat{\lambda}_R/\Gamma = d_k$ if $N^{1/2}\hat{\lambda}_M/\Gamma \in B_k$ and $|\hat{\lambda}_R| \leq 1$. If not possible, choose $\hat{\lambda}_R = \pm 1$ on the basis of the likelihood.

The resulting estimate of β is $\hat{\beta}_R$ and the estimate of the conditional median of Y given x_o is $f(\hat{\lambda}_R, x_o\hat{\beta}_R)$.

The above procedure is asymptotically the same as a restricted maximum likelihood method and is quite intuitive as it choses the point in $\mathcal{D}$ closest to $N^{1/2}\hat{\lambda}_M/\Gamma$. Note also that as N increases, the number of possible choices for scale also increases, as desired. Make the definitions:

$$q_N = N^{-1}\ \Sigma_1^N\ \tau_i^2\, x_i \rightarrow q$$

$$a_1 = [x_o\, q' - (x_o\beta)^2]/2$$

$$c_N = (-\tau_1^2/2\ \ldots\ -\tau_N^2/2)$$

$$X_N' = (x_1'\ \ldots\ x_N')$$

$$e_N = (1/4)\left\{N^{-1}\ \Sigma_1^N\ \tau_i^4 - \|q_N\|^2\right\} \rightarrow e_o > 0 .$$

<u>Theorem.</u> Using the B-D asymptotics, the limit distribution of the restricted estimator of the conditional median

$$(2.1)\qquad N^{1/2}\left[f(\hat{\lambda}_R, x_o\hat{\beta}_R) - f(\lambda = b\sigma/N^{1/2}, x_o\beta\right]$$

is given by

$$(2.2)\qquad \exp(x_o\beta)\left[\|x_o\|Z_1 + a_1\sum_k (d_k - b)I(c_o^{-1/2}Z_2 + b \in B_k)\right] ,$$

where Z_1 and Z_2 are independent standard normal random variables. The proof is in the appendix.

The Theorem shows that the estimate of the conditional median of Y given x_o based on a restricted choice of λ is not necessarily asymptotically normally distributed.

Example #1. Suppose that for any sample size we restrict our choice of $\hat{\lambda}_R$ to a fixed set, say

$$\Omega_R = \{0, \pm 1/2, \pm 1\} .$$

In this case we eventually have $\hat{\lambda}_R = 0$ so that $\mathcal{D} = \{0, \pm\infty\}$ and

$$(2.3)\qquad \text{MSE (fixed finite set)}$$

$$\rightarrow \exp(2\,x_o\,\beta)\left[\|x_o\|^2 + b^2 a_1^2\right] .$$

In simple linear regression with a symmetric design and fourth moment μ_4 satisfying (1.2), we find that at the origen $x_o = (1\ o)$, $a_1^2 = \beta_1^4/4$ and $c_o = \beta_1^4(\mu_4 - 1)/4$. In this case, while (2.3) does not serve as a very good method for predicting the individual values in Tables 1 and 2, it does, however, lead to the following qualitative conclusions, *all of which are satisfied by the simulations:*

(i) Changing the value of N from 20 to 40 while fixing Ω_R and λ basically increase b by a factor of $\sqrt{2}$. Hence, larger values of N will result in a worse performance for the RE when $\lambda \notin \Omega_R$.

(ii) Changing σ from 1 to 1/2 increases b by a factor of 2 and should result in worse performance for the RE.

(iii) Changing P_1 from 2 to 4 increases the term β_1^4 by a factor of sixteen. Such a large change should cause <u>much</u> worse performance in Table 2.

(iv) The increase in (iii) above should make the changes in the RE when one changes N or σ much more dramatic in Table 2 than in Table 1.

<u>Example #2.</u> The theory includes the MLE $\hat{\lambda}_M$ by choosing $\mathcal{D}$ dense. In this case we get

(2.4a) MSE (MLE of λ)

$$\to \exp(2x_o\beta)\left[\|x_o\|^2 + a_1^2/e_o\right] .$$

In the simple linear regression, at the origen this becomes

(2.4b) $$\exp(2\beta_o)\left[1 + (n_4-1)^{-1}\right] .$$

Note that (2.4b) is independent of the value of b .

<u>Example #3.</u> An interesting example in which the number of possible values of $\hat{\lambda}_R$ increases with N occurs when $\mathcal{D}$ = {all integers} . It is not too unreasonable to suspect that this restricted estimate will be at least comparable to the MLE, perhaps somewhat better when b=0 and hence $\lambda \in \Omega_R$, but not too much worse when b = 1/2 and $\lambda \notin \Omega_R$. In this case

(2.5) MSE (restricted procedure)

$$\to \exp(2x_o\beta)\left[\begin{array}{l}\|x_o\|^2 \\ + (a_1^2 e_o^{-1}) \sum_k (k-b)^2 \; e_o \; P\{e_o^{-1/2}Z + b \in B_k\}\end{array}\right] .$$

The only important difference between (2.4a) and (2.5) is the term

$$\sum_k (k-b)^2 e_o P\{e_o^{-1/2}Z + b \in B_k\} .$$

In Table 3 we compare the values of (2.4b) and (2.5) for the uniform simple linear regression design of the introduction with $\mu_4 = 1.79$; all comparisons are at the origen $x_o = (1\ o)$.

Table 3

Comparison of MSE for a simple linear regression design with moments $\mu_1 = \mu_3 = 0$, $\mu_2 = 1$, $\mu_4 = 1.79$, $x_o = (1\ o)$

b	β_1	$\frac{\text{MSE(MLE)}}{\text{MSE}(\lambda\text{ known})}$	$\frac{\text{MSE(RE)}}{\text{MSE}(\lambda\text{ known})}$	$\frac{\text{MSE(RE)}}{\text{MSE(MLE)}}$
0	1.5	2.27	2.37	1.04
0	2.0	2.27	2.59	1.14
0	4.0	2.27	1.03	.45
1/2	1.5	2.27	2.37	1.04
1/2	2.0	2.27	2.61	1.15
1/2	4.0	2.27	129.42	57.01

The results are somewhat surprising. First note that the case $b = 0$ corresponds to situations in which λ truely belongs to the set Ω_R. The restricted estimate does not always outperform the MLE, although it does for large β_1. What is even more interesting is the case $b = 1/2$, which is one of the simplest cases in which λ is not in the set Ω_R although it is quite close. Here we see that the restricted procedure can perform very badly indeed.

Tables 1-3 and the Theorem thus suggest that if the number of possible choices of scale is only on the order of $N^{1/2}$, the performance of the resulting estimates will differ from estimates based on the MLE of λ, in some cases being better but in others being very much worse. If one has no prior belief or evidence that only a finite number of values of λ are possible, but rather in estimating the conditional median of Y given x_o one wants to make only "reasonable" choices of λ while retaining MLE-type behavior, the number of possible choices of λ will have to be

References

Andrews, D.F. (1971): A note on the selection of data transformations. Biometrika 58, 249-254.

Atkinson, A.C. (1973): Testing transformations to normality. J. Royal Statist. Soc. Ser. B, 35, 473-479.

Bickel, P.J. and Doksum, K.A. (19): An analysis of transformations revisited. To appear in J. Amer. Statist. Assoc.

Box, G.E.P. and Cox, D.R. (1964): An analysis of transformations. J. Royal Statist. Soc. Ser. B, 26, 211-252.

Carroll, R.J. (1980): A robust method for testing transformations to achieve approximate normality. J. Royal Statist. Soc., Series B, 42, 71-78

Carroll, R.J. and Ruppert, D. (1980): Prediction and the power transformation family. To appear in Biometrika.

Efron, B. (1975): Defining the curvature of a statistical problem (with applications to second order efficiency). Ann.Statist. 3, 1189-1242.

Hájed, J. and Sidåk, Z. (1967): Theory of Rank Tests. Academic Press, New York.

Hinkley, D.V. (1975): On power transformations to symmetry. Biometrika 62, 101-111.

Roussas, G.G. (1972): Contiguity of Probability Measures. Cambridge University Press.

Appendix

We will use contiguity techniques (Hajek and Sidak (1978)). Let L_1 be the log-likelihood when $b=0$ and let L_2 be the log-likelihood for fixed $b \neq 0$. Somewhat detailed calculations show that as $N \to \infty$, $\sigma \to 0$, under the distribution L_1 with $\lambda = 0$,

$$\text{(A.1)} \qquad -(L_2 - L_1) = (b^2/8N) \sum_1^N \tau_i^4 + (bN^{-1/2}) \sum_1^N \epsilon_i \tau_i^2/2 + o_p(1) .$$

This shows that the case $b \neq 0$ is contiguous to the case $b = 0$.

Proof of the Theorem: When $\lambda = b = 0$ it follows by a Taylor expansion in λ_R that as $N \to \infty$, $\sigma \to 0$

$$\text{(A.2)} \qquad S_N = N^{1/2}\left[f(\hat{\lambda}_R, \hat{\beta}_R) - f(\lambda=0,\beta)\right] \exp(-x_o\beta)$$

$$= N^{-1/2} \sum_{i=1}^N x_o x_i' \epsilon_i + a_1 N^{1/2} \hat{\lambda}_R/\sigma + o_p(1) .$$

Also, B-D show that when $\lambda = 0$,

$$\text{(A.3)} \qquad 2e_o N^{1/2} \hat{\lambda}_M/\sigma = N^{-1/2} \sum (\tau_i^2 - x_i q')\epsilon_i + O_p(1) .$$

It is easy to check that the r.h.s. of (A.3) is asymptotically independent of the first term on the r.h.s. of (A.2). We now use the definition of $\hat{\lambda}_R$ and the convention $\sigma = r(N)\eta = r(N)$ to obtain that when $\lambda = 0$, as $N \to \infty$ and $\sigma \to 0$,

$$\text{(A.4)} \qquad S_N = N^{-1/2} \sum_{i=1}^N x_o x_i' \epsilon_i + a_1 \sum_k d_k I(N^{1/2}\hat{\lambda}_M/\sigma \in B_k) + o_p(1) .$$

We are now in a position to use Theorem 7.2 of Roussas (1972, page 38). In his notation,

$$T_N' = N^{-1/2} \sum_{i=1}^N (x_o x_i' \epsilon_i \quad q\, x_i' \epsilon_i \quad \tau_i^2 \epsilon_i)$$

$$\text{(A.5)} \qquad \Gamma = E\, T_N T_N'$$

$$h' = (0 \quad 0 \quad -b/2) .$$

One can show that the terms in (A.5) satisfy the conditions of Roussas' Theorem 7.2 so that when $\lambda = b\sigma N^{-1/2}$, as $N \to \infty$ and $\sigma \to 0$, T_N is asymptotically normally distributed with mean Γh and covariance Γ. Because of (A.3), this means that $N^{1/2}\hat{\lambda}_M/\sigma$ and the first term on the r.h.s. of (A.2) are, when $\lambda = b\sigma N^{-1/2}$, jointly asymptotically normally distributed with means $(h \quad -bx_o q/2)$, variances $(c_o^{-1}, \|x_o\|^2)$ and zero covariance. From this we obtain that (2.1) is asymptotically distributed with the same distribution as

$$\|x_o\| Z_1 - ba_1 + a_1 \sum_k d_k 1(c_o^{-1/2} Z_2 + b \in B_k) ,$$

where Z_1 and Z_2 are as in the Theorem. This completes the proof.

$\square$